

Local Structure Matching Driven by Joint-Saliency-Structure Adaptive Kernel Regression

Binjie Qin*, Zhuangming Shen, Zien Zhou, Jiawei Zhou, Jiulai Sun, Hui Zhang, Mingxing Hu and Yisong Lv

Abstract—For nonrigid image registration, matching the particular structures (or the outliers) that have missing correspondence and/or local large deformations, can be more difficult than matching the common structures with small deformations in the two images. Most existing works depend heavily on the outlier segmentation to remove the outlier effect in the registration. Moreover, these works do not handle simultaneously the missing correspondences and local large deformations. In this paper, we defined the nonrigid image registration as a local adaptive kernel regression which locally reconstruct the moving image's dense deformation vectors from the sparse deformation vectors in the multi-resolution block matching. The kernel function of the kernel regression adapts its shape and orientation to the reference image's structure to gather more deformation vector samples of the same structure for the iterative regression computation, whereby the moving image's local deformations could be compliant with the reference image's local structures. To estimate the local deformations around the outliers, we use joint saliency map that highlights the corresponding saliency structures (called Joint Saliency Structures, JSSs) in the two images to guide the dense deformation reconstruction by emphasizing those JSSs' sparse deformation vectors in the kernel regression. The experimental results demonstrate that by using local JSS adaptive kernel regression, the proposed method achieves almost the best performance in alignment of all challenging image pairs with outlier structures compared with other five state-of-the-art nonrigid registration algorithms.

Index Terms—nonrigid registration, outliers, missing correspondence, local large deformation, local model, local similarity, local structure adaptivity, kernel regression, joint saliency map

I. INTRODUCTION

Manuscript received April 19, 2013; revised September 11, 2013; This work was supported in part by NSFC (61271320, 60872102 and 60402021), NBRPC (2010CB834300), China Scholarship Council and the small animal imaging project (06-545).

*Binjie Qin is with the School of Biomedical Engineering and Med-X Research Institute, Shanghai Jiao Tong University, Shanghai, 200240, China. During 2012 to 2013, he is also with Centre for Medical Image Computing, University College London, WC1E 6BT London, U.K. (e-mail: bj Qin@sjtu.edu.cn).

Zhuangming Shen and Jiawei Zhou are with the School of Biomedical Engineering and Med-X Research Institute, Shanghai Jiao Tong University, Shanghai, 200240, China (e-mail: shenzhuangming@yahoo.com.cn; lenayui@gmail.com).

Zien Zhou is with Department of Radiology, Shanghai Renji Hospital, School of Medicine, Shanghai Jiaotong University (e-mail: zien.zhou@gmail.com).

Zhuangming Shen, Zien Zhou and Jiawei Zhou have contributed equally to this paper.

Jiulai Sun is with the Machine Vision Laboratory, University of the West of England, Bristol, BS16 1QY, U.K. (e-mail: jiulai2.sun@uwe.ac.uk).

Hui Zhang and Mingxing Hu are with Centre for Medical Image Computing, University College London, WC1E 6BT London, U.K. (e-mail: g.zhang@cs.ucl.ac.uk; mingxing.hu@ucl.ac.uk).

Yisong Lv is with the Department of Mathematics, Shanghai Jiao Tong University (e-mail: yslv@sjtu.edu.cn).

Nonrigid image registration has attracted increasing attention at motion tracking, change detection, image segmentation and surface reconstruction in the computer vision and pattern recognition for the last decade [1]–[3]. The objective of nonrigid image registration is determining the local transformations that align every structures (or features) in one (moving) image with the corresponding structures in the another (reference) image. However, owing to the image content changes over a period of time and the different physical mechanisms of multimodal imaging sensors, some local structures presented in one image appear partially or even disappear completely in another image. Such local structures without one-to-one correspondence are closely intertwined with the structures' local large deformations in the nonrigid image registration. The missing correspondences and local large deformations of structures are called outliers in this paper. At present, obtaining an accurate and robust nonrigid image registration is still a challenging task for matching these local outlier structures with missing correspondences and/or the local large deformations. In the computer vision community, the nonrigid registration with the local large deformations is also known as the large displacement optical flow problem [4].

Generally, outliers mean the extreme observations substantially different from all other ones in the real data. For nonrigid image registration, the missing correspondences and local large deformations introduce the extreme observations which appear as the extreme local geometric and intensity differences between the two images to be registered. In geometric morphology, these outliers exhibit some large and complex structural discrepancies in the locally varying spatial contexts where the underlying different local structures could deform in substantially different ways. The desired nonrigid image registration should be able to match the moving image's local structures to the corresponding reference image's structures from these various local differences. Therefore, the *local regression models* [5][6][35] that are adept at handling the locally varying differences are necessary to account for these outliers, and it provides the rationale behind this work. Recently, Gerogiannis *et al.*[7] justify that the local Bayesian regression model are favorable to model local registration transformation compared with other interpolation based registration techniques.

Over the last several decades, many relevant works were proposed to match the local structures (or features) of the two images by minimizing the differences between the two images, which include feature-based, intensity-based, and hybrid methods [1][2]. Feature-based [8][9] approaches are local model based registration methods because they always

use the local model of sparse image representation to select some corresponding features in the two images, and then directly match the local features by finding a geometrical transformation from these sparse feature correspondences. With the recent development of local feature detectors and descriptors (such as SIFT) [4][10], the feature-based methods have developed into the hybrid methods to use integrated local feature signatures [11] in characterizing each voxel in the two images. Nevertheless, the computation of registration transformation is very sensitive to the ambiguity in finding local feature correspondences in different local spatial contexts with outliers. Moreover, most feature-based methods have not solved the outlier issues that have both missing correspondences and local large deformations.

By using the complete image data to recover dense correspondences at pixel-level precision, most intensity-based nonrigid registration approaches are regarded as global model based methods that are often formulated as *global* energy minimization problems with the energy being composed of an regularization term and a similarity term [12][13]. The relative weight of similarity term and regularization term can cause the well-known trade-off between the registration accuracy and the smoothness of the deformation field [14]. In the presence of outliers, the accurate and plausible local structure matching does not exist using whole-intensity driven transformation model. The relative spatial regularization can either cause non-smooth and implausible distortion in these outlier regions or introduce over-smooth and inaccurate mapping artifacts between the whole images. This outlier problem can be solved by using a locally varying weight between regularization and image similarity [15]-[21], creating artificial correspondences [22]-[27], using cost-function masking [28]-[30], or developing SIFT flow for large displacement [4]. These approaches are largely dependent on the segmentation of outlier regions and give no full consideration to both the missing correspondences and local large deformations.

By successfully handling the locally varying differences in pattern recognition and machine learning, local kernel regression [6][36] (or nonparametric regression) is regarded as an ideal local regression model to effectively reconstruct the desired local signal while suppressing the outlier and noise effects. The normalized convolution [36] used by Suarez *et al.* [37][38] was the first application of local kernel regression in estimating dense deformation fields from sparse deformation fields. More recently, two works [39][40] also utilized kernel regression to estimate registration transformations. Xing and Qiu [42] proposed intensity-based nonparametric local smoothing to estimate local mapping transformation in a neighborhood. However, these methods do not consider the outlier problems in image registration.

To solve the outlier problem, we proposed the joint saliency map (JSM) to group the corresponding saliency structures (called Joint Saliency Structures, JSSs) in intensity-based similarity measure computation [31]. The JSM has been proved to greatly improve the accuracy and robustness of rigid [31][32] and nonrigid [10][33][34] image registration with outliers. We further think, by reflecting the local structure correspondence, the JSM also could guide the local kernel regression for accu-

rately estimating registration transformations in the nonrigid image registration with outliers.

In this paper, by integrating the JSM into the kernel regression's local adaptive estimation, we propose a new JSS adaptive kernel regression to solve the outlier problem in the non-rigid registration. Specifically, with a moving window/kernel in kernel regression based nonrigid image registration, the output dense deformation vectors are locally estimated based on the specific weights of the sparse deformation vectors within the moving window. In the presence of outliers, the weights of the sparse deformation vectors for the outliers should be as small as possible to reduce the outlier effect on giving a distorted regression of dense deformations, while the JSSs and their underlying sparse deformation vectors should be emphasized with their weights being kept as high as possible to ensure the precision of the regression computation. Furthermore, the kernel function adapts its shape [35],[44]-[47] and orientation to the reference image's local structure in order to gather more deformation vector samples of the same structure in the kernel regression, whereby the regression of local deformations can be locally compliant with the underlying local saliency structures without directing the deformation across the edges and corners of local structures.

To the best of our knowledge, this is the first work which propose the local JSS adaptive kernel regression to align the local structures by locally estimating the dense deformation fields of nonrigid image registration in the presence of outliers. An important contribution is that we use the JSM to guide local adaptive kernel regression for the accurate matching of local structures while suppressing outlier effects on the nonrigid image registration. The proposed method also makes the regression kernel not only locally adapt to the JSSs in the two images but also to the reference image's saliency structures. These two adaptivities enhance our local structure matching algorithm in achieving the accuracy and robustness of nonrigid registration of images with missing correspondences and local large deformations. The rest of this paper is organized as follows. Our algorithm is elaborated in Section 2 followed by experimental results in Section 3. The whole paper is concluded in Section 4.

II. METHODS

A. Coarse-to-fine Block Matching Scheme

The algorithm proposed by us is built upon a coarse-to-fine iterative block matching scheme similar to the one in [37][38]. The coarse-to-fine iterative framework is well known to deal with large deformation (or large displacement [4]) in nonrigid image registration [12]. Fig. 1 displays the whole coarse-to-fine iterative framework, where different levels have their own resolutions but the same procedure. At each level, the resulted global deformation is composed of initial deformation and current deformation. The initial deformation is obtained by resampling the global deformation from the previous level while the current deformation is the result of the current level involving the following two stages.

In the first stage, a discrete and sparse displacement field is created by maximizing local similarity measure for every pixel.

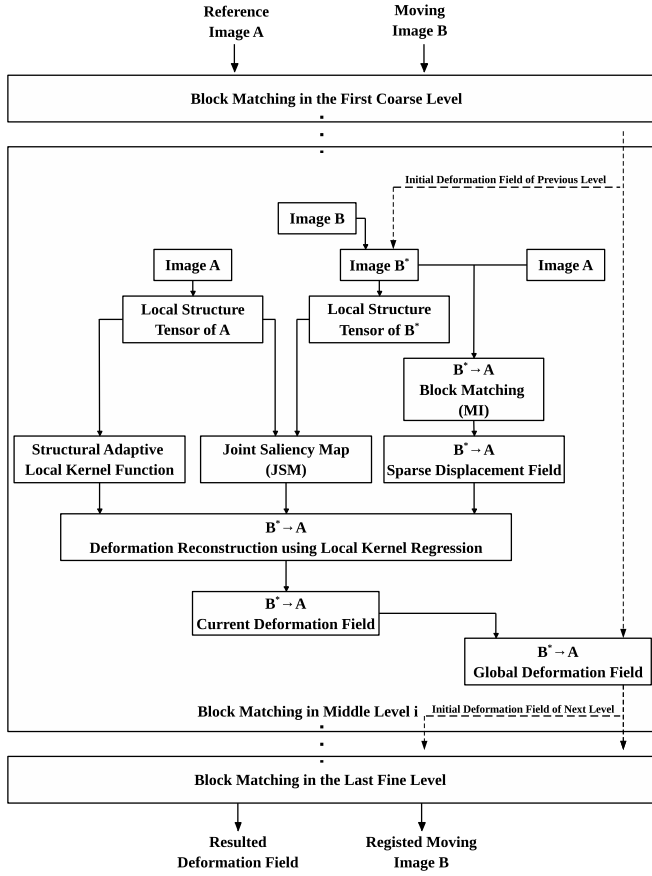


Fig. 1: Flowchart of our algorithm within coarse-to-fine framework.

The choice of local similarity measure has been studied in our previous work [42], where we employed cross-correlation (CC) instead of mutual information (MI) as the similarity measure when we matched two images at lower levels of the hierarchy so that the problem of MI's statistical consistency [43] could be solved during the hierarchical subdivision. In this study, however, we only employ MI as the local similarity measure because we have restricted the block size of the lowest level of the hierarchy. Although block matching has many advantages in obtaining the deformations of an image, implementing only this algorithm is still insufficient to avoid the irregularity of transformations such as tearing, folding and distorting. Therefore, further reconstruction constraint is indispensable to be integrated into the registration procedure. To this end, a reconstruction procedure using local kernel regression is applied to this discrete and sparse displacement fields in the second stage. Details of this stage is described in Section 2.2 to Section 2.4. After the above-mentioned two stages for each level in this coarse-to-fine framework, a smooth and dense deformation field is iteratively achieved as the global deformation at the last and finest level.

B. Kernel Regression Based Local Deformation Reconstruction

Inspired by the successful applications in modern image and video deblurring and super-resolution imaging [35],[44]-[47], we utilize kernel regression to reconstruct the dense deformation fields from the discrete and sparse displacement fields obtained through block matching. Suppose we have sparse and irregularly distributed deformation fields $\{\mathbf{y}_i, \mathbf{x}_i\}_{i=1}^P$ given in the form

$$\mathbf{y}_i = T(\mathbf{x}_i) + \varepsilon_i, \quad \mathbf{x}_i \in \Omega, \quad i = 1, \dots, P \quad (1)$$

where \mathbf{y}_i is a sparse displacement vector (response variable) at position (explanatory variable) \mathbf{x}_i , $T(\cdot)$ describes the desired dense deformation field in the moving windows (kernel) Ω with independent and identically distributed zero mean noise $\varepsilon_i = \varepsilon(\mathbf{x}_i)$. In statistics, the function $T(\cdot)$ is treated as a regression of \mathbf{y} on \mathbf{x} , $T(\mathbf{x}) = E\{\mathbf{y}|\mathbf{x}\}$. In this way, the reconstruction of nonrigid deformation fields is from the field of the regression techniques.

Suppose the point of interest \mathbf{x} to be reconstructed is near \mathbf{x}_i , then the regression of dense deformation field $T(\mathbf{x}_i)$ can be approximated by a local Taylor series expansion

$$\begin{aligned} T(\mathbf{x}_i) &\approx T(\mathbf{x}) + \{\nabla T(\mathbf{x})\}^T (\mathbf{x}_i - \mathbf{x}) \\ &\quad + \frac{1}{2} (\mathbf{x}_i - \mathbf{x})^T \{\text{Hessian}[T(\mathbf{x})]\}^T (\mathbf{x}_i - \mathbf{x}) + \dots \\ &\approx \beta_0 + \beta_1^T (\mathbf{x}_i - \mathbf{x}) \\ &\quad + \beta_2^T \text{vech}\{(\mathbf{x}_i - \mathbf{x})(\mathbf{x}_i - \mathbf{x})^T\} + \dots \end{aligned} \quad (2)$$

where $\text{vech}(\cdot)$ is a half-vectorization operator of a symmetric matrix and $\{\beta_0, \beta_1, \beta_2, \dots, \beta_N\}$ are $N + 1$ unknown parameters to be estimated. As in the 2D case, $\mathbf{x} = [x_1, x_2]^T$, we can easily estimate the unknown parameters as

$$\begin{aligned} \beta_0 &= T(\mathbf{x}) \\ \beta_1 &= \left[\frac{\partial T(\mathbf{x})}{\partial x_1}, \frac{\partial T(\mathbf{x})}{\partial x_2} \right]^T \\ \beta_2 &= \frac{1}{2} \left[\frac{\partial^2 T(\mathbf{x})}{\partial x_1^2}, \frac{\partial^2 T(\mathbf{x})}{\partial x_1 \partial x_2}, \frac{\partial^2 T(\mathbf{x})}{\partial x_2^2} \right]^T \\ &\dots \end{aligned} \quad (3)$$

Since we have known the discrete displacement vectors $\{\mathbf{y}_i\}_{i=1}^P$, we can compute $\{\beta_n\}_{n=0}^N$ by finding the optimum solution of the following weighted least squares problem

$$\min_{\{\beta_0, \beta_1, \beta_2, \dots\}} \sum_{i=1}^P [\mathbf{y}_i - \beta_0 - \beta_1^T (\mathbf{x}_i - \mathbf{x}) - \beta_2^T \text{vech}\{(\mathbf{x}_i - \mathbf{x})(\mathbf{x}_i - \mathbf{x})^T\} - \dots]^2 K_H(\mathbf{x}_i - \mathbf{x}) \quad (4)$$

where $K_H(\cdot)$ is a kernel function (see the detail in the next section), which not only smoothes the approximation but also penalizes distance away from \mathbf{x} .

In addition, if we assume $\mathbf{y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_P]^T$, $\mathbf{b} = [\beta_0, \beta_1^T, \dots, \beta_N^T]^T$, and $\mathbf{K} = \text{diag}[K_H(\mathbf{x}_1 - \mathbf{x}), K_H(\mathbf{x}_2 -$

$\mathbf{x}), \dots, K_H(\mathbf{x}_p - \mathbf{x})]$, then we can rewrite the optimization problem in a matrix form

$$\hat{\mathbf{b}} = \underset{\mathbf{b}}{\operatorname{argmin}} (\mathbf{y} - \mathbf{X}\mathbf{b})^T \mathbf{K} (\mathbf{y} - \mathbf{X}\mathbf{b}) \quad (5)$$

with

$$\mathbf{X} = \begin{pmatrix} 1 & (\mathbf{x}_1 - \mathbf{x}) & \operatorname{vech}^T\{(\mathbf{x}_1 - \mathbf{x})(\mathbf{x}_1 - \mathbf{x})^T\} & \cdots \\ 1 & (\mathbf{x}_2 - \mathbf{x}) & \operatorname{vech}^T\{(\mathbf{x}_2 - \mathbf{x})(\mathbf{x}_2 - \mathbf{x})^T\} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (\mathbf{x}_P - \mathbf{x}) & \operatorname{vech}^T\{(\mathbf{x}_P - \mathbf{x})(\mathbf{x}_P - \mathbf{x})^T\} & \cdots \end{pmatrix}$$

and the least-squares estimation solution can be expressed as

$$\hat{\mathbf{b}} = (\mathbf{X}^T \mathbf{K} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{K} \mathbf{y} \quad (6)$$

Because the zero-order Taylor series expansion known as the *Nadaray-Watson* estimator is sufficient to reconstruct the displacement vectors, the estimation of the deformation field at \mathbf{x} has the form

$$\hat{T}(\mathbf{x}) = \hat{\beta}_0 = \frac{\sum_{i=1}^P K_H(\mathbf{x}_i - \mathbf{x}) \mathbf{y}_i}{\sum_{i=1}^P K_H(\mathbf{x}_i - \mathbf{x})} \quad (7)$$

Since images have outliers, it is reasonable to consider uncertainty for each pixel [48]. Therefore, we add a weight (or certainty) function c_i to equation (7)

$$\begin{aligned} \hat{T}(\mathbf{x}) = \hat{\beta}_0 &= \frac{\sum_{i=1}^P K_H(\mathbf{x}_i - \mathbf{x}) \cdot (\mathbf{y}_i \cdot \mathbf{c}_i)}{\sum_{i=1}^P K_H(\mathbf{x}_i - \mathbf{x}) \cdot \mathbf{c}_i} \\ &= \frac{\mathbf{K} \otimes (\mathbf{y} \cdot \mathbf{c})}{\mathbf{K} \otimes \mathbf{c}} \end{aligned} \quad (8)$$

The last line of equation (8) can also be expressed in the form of normalized convolution [36], where \otimes denotes convolution operation.

Fig. 2. illustrates the discrete displacement vectors reconstructed for every pixel in tumor resection region using local kernel regression. Because block matching results inherently contain incorrect matches, which are exacerbated by the outliers in the tumor resection region. As a result, the conflicts between neighboring displacement vectors (see the several red circles shown in Fig. 2(a)) widely exist in the discrete displacement field for the tumor region. Those displacement conflicts can easily introduce the topology change of structures, such as tearing and distorting. Fortunately, all the displacement vector conflicts are removed or smoothed by the local kernel regression in Fig. 2(b), where the displacement vectors having opposite directions fully disappear with the displacement magnitudes simultaneously being smoothed. Next, to match local structures in the presence of outliers, we design structural adaptive kernel functions and robust weighing schemes for the moving windows/kernels to further boost the accuracy and robustness of the local deformation reconstruction.

C. Local structure-adaptive kernel function

As a crucial component of local kernel regression, the shape of the kernel function (or moving window) determines the spatial distribution of samples which are gathered for the quality of the locally reconstructed signal. In principle, isotropic Gaussian kernels are mostly used as kernel functions

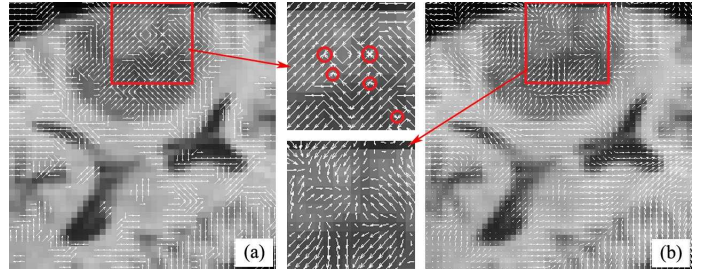


Fig. 2: **Application of kernel regression in deformation reconstruction.** (a) Discrete displacement field, (b) Reconstructed deformation field using kernel regression.

in nonparametric regression. However, traditional isotropic Gaussian kernels are insufficient to cover more samples of the same modality along some specific orientations in signal reconstruction. Besides, Gaussian kernels' fixed scales and orientations can neither detect nor enhance edge structures. These factors easily cause diffusion across lines or edges of an image. To remedy these restrictions, Pham *et al.* proposed an anisotropic kernel function to adapt its shape to the density of sampling [46]. Afterwards, Takeda *et al.* [47] utilized gradient covariance matrices to construct steering kernels, which have been proved to possess the ability to capture the edges of an image and be extremely robust to noise and perturbations of the data.

In reconstruction of dense deformation field, the local kernel function extends along local structure orientation in the reference image so that it can gather more samples of sparse deformation vectors that correspond to the same saliency structures in the reference image. Besides, the kernel function contracts in the normal orientations of local saliency structures to prevent deformation diffusion across the edges between the different structures. Based on these schemes, it can effectively reduce the risk of changing the topology of the local saliency structures in the kernel regression. Considering that local structure tensor (LST) represents the anisotropic local saliency structure information of an image[48][49], we therefore design local structure-adaptive kernel functions using the LST information in the reference image.

Before designing a local structure-adaptive kernel function at a certain pixel, we must compute LST in advance to grasp the local saliency structure around that pixel in 2D image. We assume that $I(\mathbf{x})$ denotes the intensity value at point $\mathbf{x}(x, y)$, the gradient information are expressed as $I_x = \frac{\partial I}{\partial x}$, $I_y = \frac{\partial I}{\partial y}$, and $\nabla I(\mathbf{x}) = [I_x \ I_y]^T$ indicates the local gradient vectors, then the gradient structure tensor (GST) in 2D case can be described as

$$GST(\mathbf{x}) = \nabla I(\mathbf{x}) \nabla I(\mathbf{x})^T = \begin{pmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{pmatrix} \quad (9)$$

Generally, to integrate the surrounding structural information from neighborhoods, the GST is smoothed by a Gaussian filter to derive the LST. The scale σ of the Gaussian filter is half the filter window size, namely $\sigma = 1.5$, because the size of the filter window in our experiments is 3×3 pixels (or $3 \times 3 \times 3$ voxels for the 3D image). Therefore, we derive the LST from GST and perform a principal component analysis of the LST

as follows:

$$LST(\mathbf{x}) = G_\sigma * GST(\mathbf{x}) = \lambda_u \mathbf{u} \mathbf{u}^T + \lambda_v \mathbf{v} \mathbf{v}^T \quad (10)$$

where $\{\lambda_u, \lambda_v\}$ ($\lambda_u \geq \lambda_v$) are the eigenvalues with the corresponding eigenvectors being $\{\mathbf{u}, \mathbf{v}\}$. These eigenvectors contain the information about the local orientation distribution, in which pixel values change fastest along \mathbf{u} direction while slowest along \mathbf{v} . Moreover, eigenvalues reflect both intensity variation in each direction and morphological information of a local region. For example in 2D image, $\lambda_u \approx \lambda_v \approx 0$ corresponds to a homogenous region with no measurable structure, $\lambda_u \gg \lambda_v \approx 0$ describes linear structure while $\lambda_u \geq \lambda_v \gg 0$ appears at corner structure.

With the above-mentioned LST computation in mind, we design the kernel functions to meet the following requirement: for regions containing obvious structural information such as histologic margins in a medical image, the kernels should be anisotropic with the regression computation being enhanced just along the main orientation and being suppressed along other orientations; for homogeneous regions without distinct structural information, the kernel should be isotropic with the regression computation being equal in all direction. We assume that \mathbf{x}_0 denotes a central position in 2D image, \mathbf{x} is a position in its neighborhood, $\{\mathbf{u}, \mathbf{v}\}$ are the eigenvectors of $LST(\mathbf{x}_0)$ and $\{\lambda_u, \lambda_v\}$ are the corresponding eigenvalues of $\{\mathbf{u}, \mathbf{v}\}$, thus a local structure-adaptive Gaussian kernel in 2D case is designed as

$$a(\mathbf{x}, \mathbf{x}_0) = \frac{1}{2\pi\sigma_u\sigma_v} \exp \left[- \left(\frac{\mathbf{d}_u^2}{2\sigma_u} + \frac{\mathbf{d}_v^2}{2\sigma_v} \right) \right] \quad (11)$$

$$\mathbf{d}_u = \langle \mathbf{d}, \mathbf{u} \rangle, \quad \mathbf{d}_v = \langle \mathbf{d}, \mathbf{v} \rangle, \quad \mathbf{d} = \mathbf{x} - \mathbf{x}_0$$

where \mathbf{d} is the vector from \mathbf{x}_0 to \mathbf{x} , $\{\mathbf{d}_u, \mathbf{d}_v\}$ are the projections of the vector \mathbf{d} on $\{\mathbf{u}, \mathbf{v}\}$, and the directional scales of the Gaussian kernel $\{\sigma_u, \sigma_v\}$ are determined by the anisotropy A as follows

$$\sigma_u = \frac{\alpha}{\alpha + A} \sigma_c, \quad \sigma_v = \frac{\alpha + A}{\alpha} \sigma_c \quad (12)$$

where $A = (\lambda_u - \lambda_v) / (\lambda_u + \lambda_v)$. The two directional scales of the Gaussian kernel can be adjusted by the parameter $\alpha > 0$ and the local scale σ_c . Specifically, α determines the eccentricity of the kernel while σ_c affects the number of discrete vectors that contribute to the reconstruction of continuous deformation vectors. For the sake of simplicity, the local scale is set to half the neighborhood window size for each kernel according to our experience. We also set $\alpha = 0.5$ and $\sigma_c = 1.5$ because we utilize a 3×3 pixel neighborhood window in our experiments. Similarly, the methodology of 3D LST and anisotropic kernel function computations is presented as an Appendix to this paper.

Two kernel functions for two distinct image structures are displayed at the doll images in the Fig. 3, where the crosses indicate two different kinds of typical image structures (blue cross for border and red cross for homogeneous region). In Fig. 3, the two pairs of orthogonal vectors indicate the main axes of their corresponding Gaussian kernels (blue cross at Fig. 3(c) and red cross at Fig. 3(d)). The length of the vector is determined by the scale in the direction of the vector.

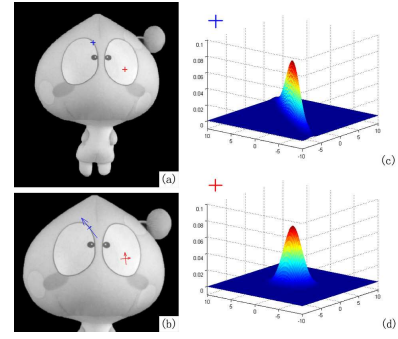


Fig. 3: **Gaussian kernels designed for different image local structures.** (a) Two labeled positions (red cross and blue one), (b) The scales and orientations of Gaussian Kernels in corresponding positions, (c) Gaussian kernel for the region with blue cross, (d) Gaussian kernel for the region with red cross.

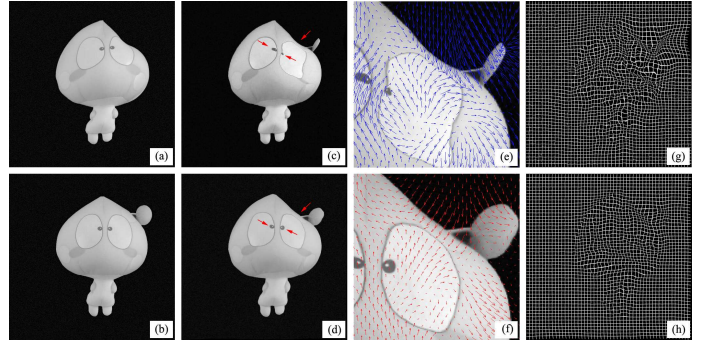


Fig. 4: **Comparison between using isotropic kernels and using local structure-adaptive kernels in kernel regression based deformation reconstruction.** (a)-(b) The reference and moving images, (c)-(d) Registered images using isotropic kernels and using local structure-adaptive kernels in kernel regression, (e)-(f) Local displacement vector fields for (c)-(d), (g)-(h) Global deformation mesh fields for (c)-(d).

Fig. 4 illustratively explain why we prefer local structure-adaptive kernel to isotropic kernel in our structure-adaptive kernel regression. The regions pointed by red arrows are small scale structures which have local large deformations. The directions of the displacement vectors (spaced every 5 pixels) in these small structures are inevitably conflicted with those of the large deformations of the neighboring structures. These deformation conflicts that are introduced by the opposite displacement vectors can easily cause tearing, folding or distorting of the local small scale structures. For example, the eyes in Fig. 4(c) display distortion owing to the conflict of the deformation directions displayed in Fig. 4(e). Comparatively, our local structure-adaptive kernels suppress the contributions of the displacement vectors which are not consistent in the structure orientation, the deformation conflicts are therefore removed in Fig. 4(f) with no distortion existing in the eyes at Fig. 4(d).

Fig. 4(h) demonstrates the overview of mesh deformation (10 pixels vertex spacing) in the local structure-adaptive kernel regression, which can produce the smooth adaptivity of local meshes deformation to the local anisotropic structures. Obviously, this local structure-adaptive kernel regression obtain smooth mesh boundaries which are consistent with the local

structures' boundaries. However, isotropic kernel in kernel regression easily produce irregularly deformed local meshes which are not adaptive to the local structures, so that it is very difficult to identify boundaries of local structures from the non-smooth mesh deformation in Fig. 4(g).

D. Robust Weight Mechanism using JSM

In original kernel regression, the weight $c(\mathbf{x})$, between 0 and 1, specifies the reliability (or certainty) at \mathbf{x} for the local estimation in a moving window/kernel. In modern local regression model, the weight function always describes the spatial dependence in the locally varying special context. With locally data-dependent weights, recently popularized and very effective image filter are developed in image and video processing field [6]. For example, the Gaussian function of a residual error with an acceptable range of error σ_r [46] severs as a new weight function in the neighborhood of \mathbf{x}_0

$$c(\mathbf{x}, \mathbf{x}_0) = \exp\left(-\frac{|I(\mathbf{x}) - \hat{I}(\mathbf{x}, \mathbf{x}_0)|^2}{2\sigma_r^2}\right) \quad (13)$$

where $I(\mathbf{x})$ denotes a measured intensity at position \mathbf{x} , and $\hat{I}(\mathbf{x}, \mathbf{x}_0)$ is an estimated intensity at \mathbf{x} using an initial polynomial expansion at \mathbf{x}_0 .

Since the regions with salient structure information have real influence over locally adaptive image processing based on kernel regression, E. Suarez *et al.* [37] proposed a weight scheme in the kernel regression of registration transformations by using the scalar measure of a local structure in reference image. However, for nonrigid image registration, the salient structural regions in reference image may introduce non-corresponding salient structural regions at same locations in moving image. Therefore, the method proposed by E. Suarez *et al.* [37] is most likely to assign wrong high weights to the salient structures that have missing correspondences.

To avoid the above-mentioned mis-assignment and minimize the outlier impact on the reconstruction of deformation field, we propose a robust weight mechanism by simultaneously considering the matching degree of local salient structures in the overlapping parts of the two images. In our previous work [31], it has been proved that the application of JSM is effective in tackling registration problems with outliers by emphatically grouping the JSSs into intensity-based similarity measure. Continuing the success of JSM, we further deploy the concept of JSM into our robust weight mechanism and pay more attention to the JSSs in the two images. Those JSSs and their incurring deformation should be emphatically treated in the local kernel regression to reconstruct the dense deformation fields from the sparse deformation fields that contain outliers.

With the general JSM-based weight scheme in mind, we should first construct a saliency map to indicate the local salient structure distribution in each image. Since LSTs mentioned in Section 2.3 contain sufficient structural information in a region, we can reasonably use them to reflect the saliency map of an image. Inspired by the center-surround mechanism [50] which has defined the intensity-contrast-based visual saliency map, we define a saliency operator based

on the contrast among neighboring structure tensors. This contrast emphasizes the dissimilarity or discrepancy between neighboring local structure tensors. Specifically, for a given point \mathbf{x}_0 and its neighborhood Ω , the saliency value $S(\mathbf{x}_0)$ at \mathbf{x}_0 in a salient map can be computed through

$$S(\mathbf{x}_0) = \text{avg}_{\mathbf{x} \in \Omega} \|LST(\mathbf{x}) - LST(\mathbf{x}_0)\|_D \quad (14)$$

where $\|\cdot\|_D$ defines a distance metric describing the dissimilarity between two LSTs, which is detailed in the following section. The operator avg computes the average of the dissimilarities within the neighborhood Ω of \mathbf{x}_0 . Traditional tensor similarity measures such as fractional anisotropy (FA) and cosine similarity measure are not appropriate for defining tensor-based saliency operator because they only compare either scales or orientations of two tensors. Fortunately, some improved tensor similarity measure computing both scale information and orientation information have been reported. In [51], H. Zhang *et al.* introduced diffusion tensor metric, which is defined as

$$\|T_1 - T_2\|_L = \sqrt{\frac{8\pi}{15}(\|T_1 - T_2\|_C^2 + \frac{1}{2}Tr^2(T_1 - T_2))} \quad (15)$$

where $\|T_1 - T_2\|_C = \sqrt{Tr(T_1 - T_2)^2}$ is the Euclidean distance between two tensors $\{T_1, T_2\}$, Tr is the operator for computing the trace of matrix. Afterwards, H. Zhang *et al.* [52] modified equation (15) to only focus on the anisotropic components between two tensors. The modified equation that is adopted at equation (14) can be expressed as

$$\|T_1 - T_2\|_D = \sqrt{\frac{8\pi}{15}(\|T_1 - T_2\|_C^2 - \frac{1}{3}Tr^2(T_1 - T_2))} \quad (16)$$

In a saliency map, the saliency values are general representation of the local structure distribution in an image. Low saliency values always appear in the homogeneous and background regions while high saliency values are assigned to the edge regions of saliency structures owing to the highlighted contrast among neighboring LSTs in these regions. After the two normalized salient maps were achieved to indicate the local structure distribution, JSM is ready to describe the matching degree between the two saliency maps at every pixel pair in the overlapping regions of the two images. Given a point \mathbf{x}_R in the reference image and its corresponding point \mathbf{x}_M in the moving image after initial transformation, their joint-saliency value in a JSM is defined as

$$JS(\mathbf{x}_R, \mathbf{x}_M) = \min\{S_R(\mathbf{x}_R), S_M(\mathbf{x}_M)\} \frac{A \cdot B}{B + \|LST(\mathbf{x}_R) - LST(\mathbf{x}_M)\|_D} \quad (17)$$

where $\{S_R(\cdot), S_M(\cdot)\}$ denote the saliency values in the saliency maps of the reference and the moving images. The empirical parameter A and B are used to normalize the JSM values into a final value between 0 and 1. In our experiments, $A = 10$ and $B = \frac{1}{2} \max(\|LST(\mathbf{x}_R) - LST(\mathbf{x}_M)\|_D)$. It should be noted that it may introduce a situation that both of the corresponding pixels are assigned high saliency values in the saliency maps, while their local variations of gradient

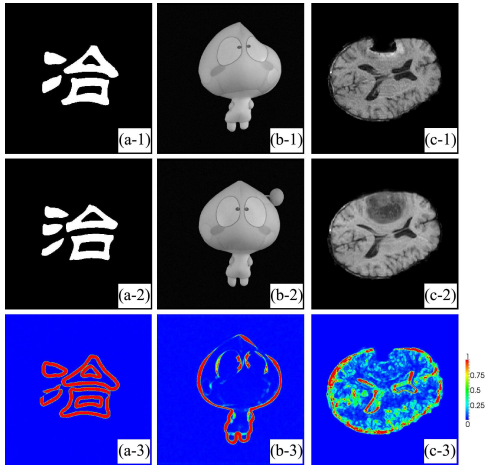


Fig. 5: JSM Examples with colour scale representing different joint saliency values. Each column shows one case. The reference images and the moving ones are shown at the top and middle rows. Their JSMs are displayed at the bottom rows where the red regions correspond to the higher joint saliency values while the blue regions correspond to the lower joint saliency values.

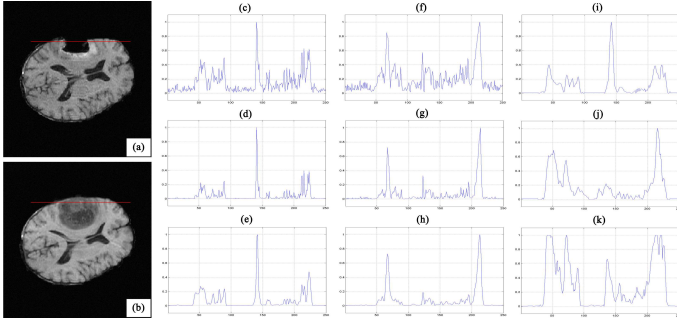


Fig. 6: The reference and the moving images and their gradient magnitude, largest eigenvalue profiles for GST and LST, and the JSM magnitude. (a)-(b) The reference and moving images, (c)-(e) Gradient magnitude profiles, largest eigenvalue profiles of GST and largest eigenvalue profiles of LST of the red line in (a), (f)-(h) Gradient magnitude profiles, largest eigenvalue profiles of GST and largest eigenvalue profiles of LST of the red line in (b), (i)-(j) Saliency value profiles of the red lines in (a)-(b), (k) JSM value profiles of the red lines in (a)-(b).

orientations are in fact totally different. To avoid this situation, we also consider the dissimilarity measure between $LST(\mathbf{x}_R)$ and $LST(\mathbf{x}_M)$ at the denominator in equation (17).

Fig. 5 shows some examples of normalized JSM with the colour scale representing different joint saliency values. The high joint saliency values being represented by red colour suggest that the underlying pixel pairs come from the JSSs. On the contrary, the regions with low JSM values are rendered in blue colour, which indicates that the underlying pixel pair originates from either homogeneous regions or outlier regions. The discrete displacement vectors in these red JSS regions are expected to contribute more to the kernel regression than the blue regions having low JSM values, this scheme is therefore called JSS adaptive kernel regression for nonrigid image registration.

The JSM in our study mainly responds to the corresponding high-gradient edge pixels. However, it does not simply high-

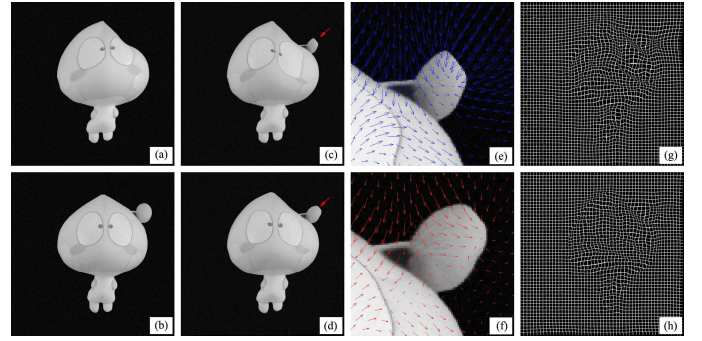


Fig. 7: Comparison between using JSM-based robust weight mechanism and not using it in our method. (a)-(b) The reference and moving image. (c)-(d) Registered images without and with using a JSM-based robust weight mechanism, (e)-(f) Local displacement vector fields of (c)-(d), (g)-(h) Global deformation mesh of (c)-(d).

light the common image gradients in the two images. Fig. 6 presents the image gradient magnitude, the largest eigenvalues of GSTs and LSTs, the saliency values and the JSM values profiles of the same red line at the two images (Fig. 6(a)-(b)). For easy comparison, the range of ordinates in Fig. 6(c)-(k) are bound to $[0,1]$. As shown in Fig. 6, the image gradient features in Fig. 6(c) and Fig 6(f) are very sensitive to noise and do not agree with each other at each overlapping location. The noise sensitivity are gradually reduced by using the GSTs (Fig. 6(d) and Fig. 6(g)) and the LSTs (Fig. 6(e) and Fig. 6(h)). The saliency values of the two images in Fig. 6(i)-(j) are robust to noise due to their computing the regional contrast of LSTs through equation (14). Moreover, the structural image information in a large region is also comprehensively considered according to equation (14). As a result, the JSM values (Fig. 6(k)) computed through the saliency values can accurately preserve the JSSs in larger capture range with smaller variability than the image gradients.

Because of the outliers introduced by missing correspondence, local large deformation and incorrect block matching, the dense deformation fields can not be interpolated from the sparse displacement vectors in block matching. The JSS adaptive kernel regression is then used to reconstruct the dense deformation fields from the sparse displacement vectors, i.e., smooth the impact of outliers on the deformation reconstruction. Due to the expected deformation in the outlier region being consistent with its neighboring deformations, the JSM in the neighboring regions is used to assign different weights to the different displacements of the neighboring structures, only those neighboring deformations with high JSM values indicating the consistency in structure orientations are given high weights in kernel regression based deformation reconstruction.

Fig. 7 illustrates the improvements on the deformation field reconstruction after introducing the JSM-based robust weight mechanism in the local JSS adaptive kernel regression. The regions pointed by red arrows in Fig. 7(c)-(d) are outlier regions with missing correspondences and local large deformations. Without JSM-based robust weight mechanism, the converged displacement vectors (5 pixels spacing) from conflicting directions (See Fig. 7(e)) in the outlier region

spread the distortion effect into the eye region (see Fig. 7(c)). Due to the JSM-based robust weight mechanism introducing weighted smoothing effect on the magnitude and direction of displacement vectors (see Fig. 7(f)), the moving image's eye distortion is removed with the outlier region's structure deformations being simultaneously matched to those of the reference image (see Fig. 7(f)). Compared with the deformation meshes (10 pixels vertex spacing) in Fig. 7(g), the deformation meshes at Fig. 7(h) display the overall smoothness improvement for the structural deformations at the outlier regions.

III. EXPERIMENTAL RESULTS

In comparison with several state-of-the-art nonrigid registration approaches to validate the proposed algorithm on some challenging images with missing correspondences and local large deformations, we choose Advanced Normalized Tools (ANTs)¹ with elastic transformation and MI (AMI) [53] due to the ANTs being placed the first in the EMPIRE10 challenge [56] evaluating 34 total registration algorithms. We also include Diffeomorphic Demons with Diffusion-like regularization (DDD)² [54], fast B-Spline with MI (BMI)³, AMI with cost-function Masking (AMM) and Large displacement Optical Flow (LOF) algorithms⁴ [4]. The parameters of our methods are: the number of pyramid levels is 5; the local similarity measure is mutual information. We set the parameters of AMI and AMM as: the histogram bin size is 32; the number of pyramid levels is 3; the iterations are set to $100 \times 100 \times 10$; the gradient step is 10; the default regularization is Gaussian filtering with a sigma of 3. The parameters of DDD method are set as follows: the variance of smoothing kernels is 2; the step scale is 1; the number of pyramid levels is 5; the number of iterations is 100. The parameters of the BMI method are selected as: the number of iterations is set to 100; the grid size is 15; the histogram bin size is 32; the spatial sample is 50000; the maximum deformation is 20. We choose the default parameters of the LOF method as: the tuning parameters for regularity constraint, the point correspondence constraint and the constraint on the gradient are 30, 300 and 5; downsampling factor is 0.95; the numbers of outer iterations and inner iterations are set to 10 and 5. With those parameters all the methods mentioned above achieve their best performances.

To evaluate the performance of the six competing methods, both registration accuracy and efficiency are estimated during the assessment. Validating the registration accuracy of binary image pairs is easy to recognize the badly-aligned regions by using the difference image between the reference and the registered moving images. However, the evaluation based on the difference images is not reliable for grayscale image registration [55]. Due to the registration errors measured at densely distributed landmarks being considered as the standard for evaluating registration accuracy of grayscale images [55], we manually selected a large number of appropriate landmark

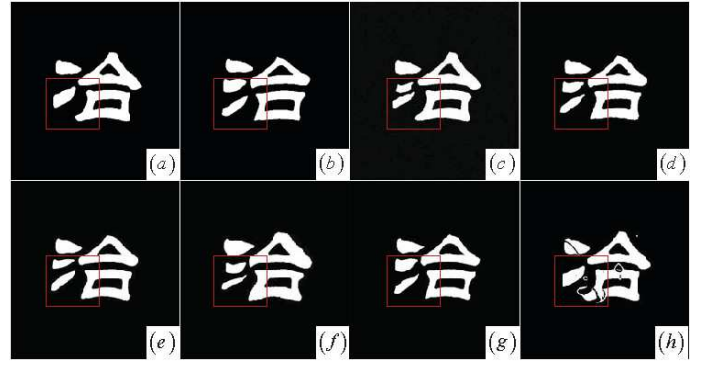


Fig. 8: **Chinese character image registration.** (a)-(b) The reference and moving images (c) the proposed method, (d) AMI, (e) DDD, (f) BMI, (g) AMM, (h) LOF.

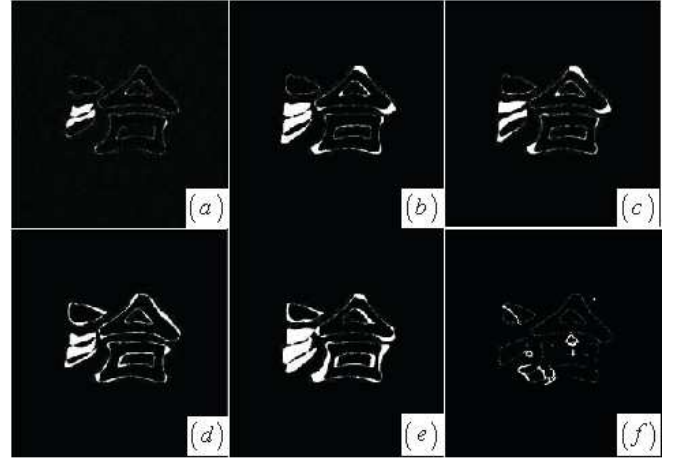


Fig. 9: **Difference images from the six Chinese character registration results.** (a) the proposed method, (b) AMI, (c) DDD, (d) BMI, (e) AMM, (f) LOF.

pairs in the reference and the registered moving images to accurately evaluate the registration accuracy of the grayscale images. The selected landmarks fully exclude outlier features but identify some corresponding small scale saliency structures (having local large deformations) around the outlier regions. Therefore, the matching accuracies of six registration methods for the local structures with outliers are fully assessed by using the Mean Registration Error (MRE) and Standard Deviation (SD) in pixels between these selected landmarks.

The reference and the moving binary images in the first experiment are two similar Chinese characters at Fig. 8(a)-(b). The two strokes in the left part of the reference image correspond to the upper left and lower left ones in the moving images, while the middle stroke (outlier) in the left part of moving image has missing correspondence (see the red rectangles in the Fig. 8). Moreover, the triangular and the rectangular openings at the right part of the character are narrowed. The local large deformations are apparent at the lower left corner of the rectangular region and the low right corner of the triangular region. Besides, the lower left stroke from southwest to northeast is lengthened with counter-clockwise rotation. Fig. 8(c)-(g) are the registered results of our approach, AMI, DDD, BMI, AMM and LOF. The strokes

¹<http://www.picsl.upenn.edu/ANTs>

²<http://mipav.cit.nih.gov>

³<http://www.slicer.org>

⁴<http://www.seas.upenn.edu/~katet/LDOF.html>

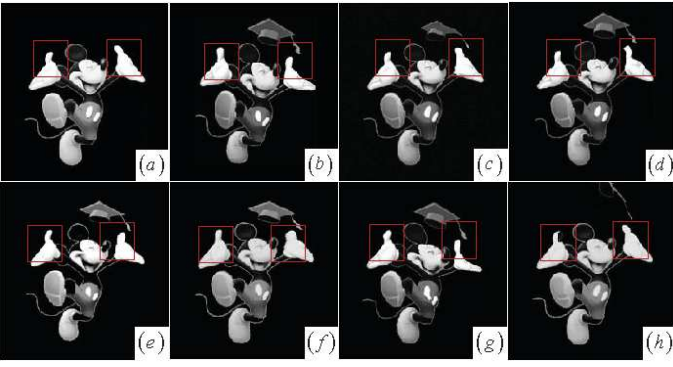


Fig. 10: **Mickey image registration.** (a)-(b) The reference and moving images, (c) the proposed method, (d) AMI, (e) DDD, (f) BMI, (g) AMM, (h) LOF.

at the left part and the corner regions of the two openings are locally deformed more or less in the six registered images. Though the constrained cost-function masking has masked the middle stroke in moving image, AMM method perform poorly in matching local structures with local large deformations in Fig. 8(g). The LOF method changed the topology of moving image by fully removing the middle stroke in the left part and introducing some artifacts in the character at Fig. 8(h). Overall, the proposed method (see Fig. 8(c)) has produced the best structure deformation among the six registered results.

The performance of our proposed approach could be clearly validated from the difference images (see Fig. 9) of the six registration results, where the white regions represent the discrepancies between the reference image and the registered moving image. Although the difference image of LOF method has smallest white regions among the six methods, the LOF method also introduce distinct artifacts and topology change in the registered Chinese character. Without robust scheme tackling missing correspondence and local large deformation simultaneously, the AMM method (Fig. 9(e)) even performs a little worse than the AMI method (Fig. 9(b)). The white regions in the difference image (Fig. 9(a)) of our approach are the least among the other five results which preserve the topology of the middle stroke. It validates that our proposed method match well the moving image's local structures into the corresponding structures at the reference image.

The second experiment involves aligning two grayscale Mickey images, which have the outlier doctoral hat shown in the moving image (Fig. 10(b)). Moreover, the local structure's large deformations occur in Mickey's left thumb, left hand, right thumb (see the red rectangles in the Fig. 10), right shoe and right button in Mickey's belly. Consequently, validating the performance of registration methods is largely dependent on the deformed results of these structures. Fig. 10(c)-(h) show the registered results of the six methods, where our proposed approach outperforms the other five methods by perfectly deforming those local structures to desired positions. In contrast, the morphologies of Mickey's left hand in Fig. 10(d) and (g) are changed by AMI and AMM methods. The Mickey's left thumb, right shoe and right button in Fig. 10(e) have not changed by DDD method. The Mickey's left thumb has not deformed and the left palm become thicker in Fig.

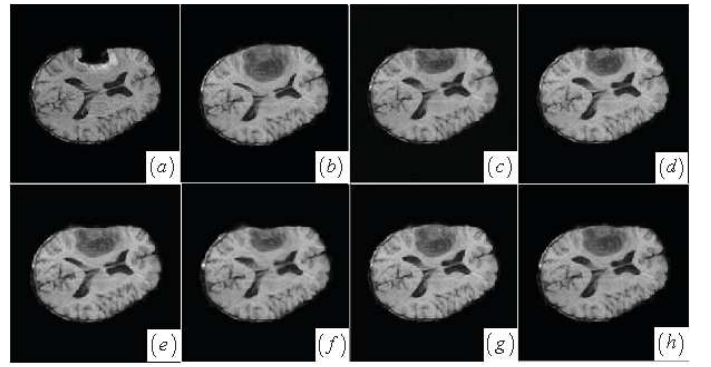


Fig. 11: **Brain tumor resection image registration.** (a)-(b) The reference and moving images, (c) the proposed method, (d) AMI, (e) DDD, (f) BMI, (g) AMM, (h) LOF.

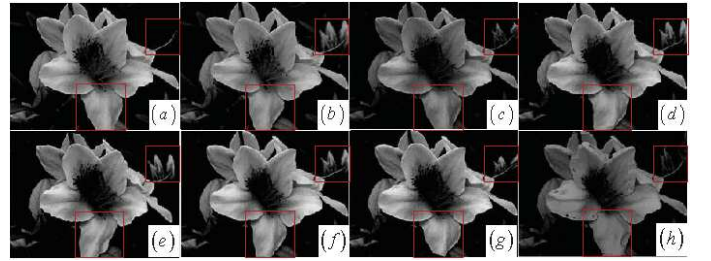


Fig. 12: **Flower image registration.** (a)-(b) The reference and moving images, (c) the proposed method, (d) AMI, (e) DDD, (f) BMI, (g) AMM, (h) LOF.

10(f) after BMI registration. In this case, though with the constrained cost-function masking for the doctoral hat, AMM method has no improvement in matching local structures (Fig. 10(g)). The LOF method introduces an undesired artifact in the right thumb of the Mickey (see the left red rectangle in the Fig. 10(h)).

Another grayscale image registration is matching pre- and post-operative brain tumor images. Brain tissue severely suppressed by tumor in the preoperative image (Fig. 11(b)) expands after tumor resection. Tumor resection not only brings missing correspondences into the tumor region of the post-operative image (Fig. 11(a)) but also incurs local large deformations that are caused by brain shift. A successful registered result of this case should properly deform pre-operative brain tissue according to the post-operative image regardless of tumor resection. Visual inspection has revealed that the proposed, AMI, AMM and LOF methods (see Fig. 11(c)-(d),(g)) apparently perform better than the DDD and BMI methods (see Fig. 11(e)-(f)) because the local brain deformation resulted from the latter two methods is either insufficient or somewhat excessive. This visual valuation is further confirmed by validating landmark-based registration error in the following section.

A more challenging flower image registration is displayed in Fig. 12, where the small scale stamen filament in the right part of the image is largely deformed while some buds behind the stamen filament of the moving image being disappeared in the reference image (see the top red rectangles in the Fig. 12). Matching small scale structures that have large

TABLE I: Registration errors (Mean+SD) of the six methods for the three grayscale image registrations

proposed	AMI	DDD	BMI	AMM	LOF
1.27±3.09	1.56±3.38	6.07±5.93	3.23±5.51	12.03±14.80	3.55±3.21
0.97±1.91	0.95±1.48	1.60±3.08	1.15±1.89	1.16±1.99	1.97±1.93
1.14±2.96	1.69±3.49	8.38±8.82	4.42±5.65	1.55±3.49	2.72±2.35

deformations and missing correspondence is very difficult for nonrigid image registration. In the case of Fig. 12, an ideal registration algorithm should accurately match not only the large petal (see the bottom red rectangles) but also the small scale stamen filament while simultaneously keeping reasonable local deformation consistency in the spatial context. Among these six methods, only the proposed approach aligns not only the small scale stamen filament but also the large scale petal by computing their reasonable and accurate deformations. The AMI (Fig. 12(d)) and AMM (Fig. 12(g)) methods achieve better registration performances than the DDD and BMI methods (Fig. 12(e)-(f)). The LOF (Fig. 12(h)) method gets good deformations for most structures, but also apparently intrude some artifacts in the petals of the registered moving image.

The MRE and SD of the manually selected landmarks for the six methods in the three grayscale image registration are listed in Tab. 1. The proposed method for the Mickey image achieves the smallest registration error of 1.27 ± 3.09 pixels while the registration errors of AMI, DDD, BMI, AMM and LOF methods are greater than or equal to 1.56 ± 3.38 pixels. Compared with other method, the proposed method and AMI have achieved sub-pixel registration accuracy for the brain tumor resection images with the registration errors of 0.97 ± 1.91 and 0.95 ± 1.48 pixels, respectively. As for the flower image, the proposed method gets the smallest registration error of 1.14 ± 2.96 pixels, while the registration errors of other five methods are greater than or equal to 1.69 ± 3.49 pixels. In average, the proposed method maintains almost the best performance in comparison with other five methods. Although the original AMI method in the brain tumor resection image registration has a slight advantage over the proposed method, using cost-function masking makes AMM method worse than the proposed method in matching locally deformed structures. Simply setting the brain-tumor-region's value to zero by cost-function masking is not enough to accurately match local salient structures with missing correspondences and local large deformations. Due to the LOF method only considering the effect of large displacements, its performance is not desired for the nonrigid registration with both missing correspondences and local large deformations.

Table 2 lists the computation time needed for the six algorithms at the four image registration, where the image resolution of case 1-3 is 372×392 pixels, and that of case 4 is 384×288 pixels. All the six methods are operated on a PC of Pentium(R) Dual-Core 3.20 GHz CPU with 2 GB memory.

TABLE II: Computation runtime in seconds for the six registration methods (Pentium(R) Dual-Core 3.20 GHz RAM 2.0 GB)

Cases	proposed	AMI	DDD	BMI	AMM	LOF
1	48	38	12	31	38	112
2	49	56	13	42	45	144
3	48	21	12	43	20	113
4	36	21	10	75	7	168

IV. CONCLUSION

In this paper, considering the outlier structures that have missing correspondences and local large deformations in non-rigid image registration, we use JSS adaptive kernel regression to reconstruct dense deformation field (for the moving image) from the sparse deformation field hierarchically computed from multi-resolution block matching. Specifically, the proposed local JSS adaptive kernel regression implements two local adaptivities into the underlying saliency structures in the reference images and the JSSs in the two images to accurately and robustly match corresponding local structures with missing correspondence and local large deformation.

Nevertheless, there are still some issues that need to be further addressed in our future work. First, the local scale mentioned in Section 2.3 is set as a constant value for the sake of simplicity at each anisotropic Gaussian kernel. As shown in some registered results, however, the deformations of some small scale structures (such as the eyes in the doll image at Fig. 8) are affected somewhat badly by the deformations of the surrounding large scale structures. This situation is caused by the unchangeable local scales. Indeed, the local scale (the width of kernel) controls the number of discrete displacement vectors contributing to the reconstruction of dense deformation vectors from sparse displacement vectors. Therefore, the choice of local scale significantly affects the final registration result. A self-adjustable local scale according to the local structure properties is expected to automatically adjust the number of discrete displacement vectors participating in the deformation reconstruction. To our knowledge, almost no attention has been paid on the self-adjustable local scale for nonrigid image registration during the last decade.

Second, we could also improve the proposed algorithm by replacing the block matching with other feature-based or hybrid nonrigid registration algorithms so that we can accurately initialize deformation field estimation and apply the proposed method to match multimodal images. As for the diffeomorphism of nonrigid image registration [53][54][56], the missing correspondences and local large deformations make it unrealistic for nonrigid registration method enforcing the diffeomorphic local structure matching. However, we could initialize our method with some diffeomorphic registration algorithms to find the local structures' optimal diffeomorphic matching for the subsequent JSS adaptive kernel regression. At last, further researches are required to reduce the computational cost of our approach. At present, more than half of the total running time for our proposed method is spent on the local adaptive Gaussian kernel reconstruction and subsequent adaptive kernel regression at every pixel. We should design fast method [57] to estimate discrete local structure-adaptive Gaussian kernel and implement structural adaptive

kernel regression at every pixel. All these improvements could prompt us to make further contributions to the nonrigid image registration.

APPENDIX A

3D LOCAL STRUCTURE TENSOR AND ANISOTROPIC KERNEL FUNCTION

The 3D LST can be computed for every point $\mathbf{x}(x, y, z)$ in 3D image as follows. With the gradient being computed as $I_x = \frac{\partial I}{\partial x}$, $I_y = \frac{\partial I}{\partial y}$, $I_z = \frac{\partial I}{\partial z}$ and the $\nabla I(\mathbf{x}) = [I_x \ I_y \ I_z]^T$ indicating the local gradient vectors, the 3D GST is estimated as

$$GST(\mathbf{x}) = \nabla I(\mathbf{x})\nabla I(\mathbf{x})^T = \begin{pmatrix} I_x^2 & I_x I_y & I_x I_z \\ I_x I_y & I_y^2 & I_y I_z \\ I_x I_z & I_y I_z & I_z^2 \end{pmatrix} \quad (18)$$

while the smoothed 3D LST being

$$LST(\mathbf{x}) = G_\sigma * GST(\mathbf{x}) = \lambda_u \mathbf{u}\mathbf{u}^T + \lambda_v \mathbf{v}\mathbf{v}^T + \lambda_w \mathbf{w}\mathbf{w}^T \quad (19)$$

where $\{\lambda_u, \lambda_v, \lambda_w\}$ ($\lambda_u \geq \lambda_v \geq \lambda_w$) are the eigenvalues and their corresponding eigenvectors are denoted as $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$. The voxel values change fastest along \mathbf{u} direction while slowest along \mathbf{w} direction.

To derive the 3D kernel function at \mathbf{x}_0 , we express the anisotropic 3D Gaussian function $a(\mathbf{x}, \mathbf{x}_0)$ using the eigenvalues and corresponding eigenvectors of 3D LST.

$$a(\mathbf{x}, \mathbf{x}_0) = \frac{1}{\sqrt{8\pi^3} \sigma_u \sigma_v \sigma_w} \exp \left[- \left(\frac{\mathbf{d}_u^2}{2\sigma_u} + \frac{\mathbf{d}_v^2}{2\sigma_v} + \frac{\mathbf{d}_w^2}{2\sigma_w} \right) \right] \quad (20)$$

$\mathbf{d}_u = \langle \mathbf{d}, \mathbf{u} \rangle, \ \mathbf{d}_v = \langle \mathbf{d}, \mathbf{v} \rangle, \ \mathbf{d}_w = \langle \mathbf{d}, \mathbf{w} \rangle, \ \mathbf{d} = \mathbf{x} - \mathbf{x}_0$

where \mathbf{d} is the vector from \mathbf{x}_0 to \mathbf{x} , $\{\mathbf{d}_u, \mathbf{d}_v, \mathbf{d}_w\}$ are the projections of the vector \mathbf{d} on $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$. In order to estimate the directional scales of the 3D anisotropic Gaussian kernel [58], we first compute the anisotropy $\{a_{vw}, a_{uw}\}$ of 3D LST with

$$a_{vw} = \frac{\lambda_v - \lambda_w}{\lambda_u + \lambda_v + \lambda_w}, \quad a_{uw} = \frac{\lambda_u - \lambda_w}{\lambda_u + \lambda_v + \lambda_w} \quad (21)$$

We further estimate the spatial dependent corner strength $C(\mathbf{x}_0)$ as

$$C(\mathbf{x}_0) = (1 - a_{vw} - a_{uw}) |\nabla I(\mathbf{x}_0)|^2 \quad (22)$$

where $|\nabla I(\mathbf{x}_0)|^2$ is the local gradient strength at the point \mathbf{x}_0 . Therefore, we obtain the three directional scales of the 3D Gaussian kernel as

$$\begin{aligned} \sigma_u &= \frac{\sigma_c(1 - a_{vw} - a_{uw})}{1 + C(\mathbf{x}_0)} \\ \sigma_v &= \frac{\sigma_c(1 - 2a_{vw})}{1 + C(\mathbf{x}_0)}, \quad \sigma_w = \frac{\sigma_c}{1 + C(\mathbf{x}_0)} \end{aligned} \quad (23)$$

ACKNOWLEDGMENT

The authors would like to thank Simon K. Warfield for providing MR brain images, E. Suarez-Santana for opening the source code in ITK project and Katerina Fragkiadaki for opening the source code of large displacement optical flow. The authors thank the open source ANTs project, MIPAV, Diffeomorphic Demons, 3D Slicer and ITK project.

REFERENCES

- [1] R. Liao, L. Zhang, Y. Sun, S. Miao, C. Chef d'hotel, *A Review of Recent Advances in Registration Techniques Applied to Minimally Invasive Therapy*, IEEE Trans. Multimedia, 2013, in publication, doi:10.1109/TMM.2013.2244869
- [2] I.-H. Kim, Y.-C. M. Chen, D. L. Spector, R. Eils, K. Rohr, *Nonrigid Registration of 2-D and 3-D Dynamic Cell Nuclei Images for Improved Classification of Subcellular Particle Motion*, IEEE Trans. Imag. Proc., vol. 20, no. 4, pp. 1011-1022, 2011.
- [3] S. Faisan, N. Passat, V. Noblet, R. Chabrier, C. Meyer, *Topology Preserving Warping of 3-D Binary Images According to Continuous One-to-One Mappings*, IEEE Trans. Imag. Proc., vol. 20, no. 8, pp. 2135-2145, 2011.
- [4] T. Brox and J. Malik, *Large Displacement Optical Flow: Descriptor Matching in Variational Motion Estimation*, IEEE Trans. Pattern Anal. Mach. Intel., vol. 33, no. 3, pp. 500-513, 2011.
- [5] C. D. Lloyd, *Local Models for Spatial Analysis*, Second Edition, CRC Press, Boca Raton, 2011.
- [6] P. Milanfar, *A Tour of Modern Image Filtering: New insights and methods, both practical and theoretical*, IEEE Signal Processing Magazine, 30(1): 106-128, 2013.
- [7] D. Gerogiannis, C. Nikou, A. Likas, *Registering sets of points using Bayesian regression*, Neurocomputing, 89, 122-133, 2012.
- [8] M. Izadi and P. Saeedi, *Robust Weighted Graph Transformation Matching for Rigid and Nonrigid Image Registration*, IEEE Trans. Imag. Proc., vol. 21, no. 10, pp. 4369-4382, 2012.
- [9] G. Auzias, O. Colliot, J. A. Glaunes, M. Perrot, J.-F. Mangin, A. Trouve, and S. Baillet, *Diffeomorphic Brain Registration Under Exhaustive Sulcal Constraints*, IEEE Trans. Med. Imag., vol. 30, no. 6, pp. 1214-1227, 2011.
- [10] Z. Gu and B. Qin, *Nonrigid Registration of Brain Tumor Resection MR Images Based on Joint Saliency Map and keypoint Clustering*, Sensors, vol. 9, no. 12, pp. 10270-10290, 2009.
- [11] S. Liao and A. C. S. Chung, *Nonrigid Brain MR Image Registration Using Uniform Spherical Region Descriptor*, IEEE Trans. Image Processing, vol. 21, no. 1, pp. 157-169, 2012.
- [12] P. Hellier, C. Barillot, E. Memin, and P. Perez, *Hierarchical estimation of a dense deformation field for 3-D robust registration*, IEEE Trans. Med. Imag., vol. 20, no. 5, pp. 388-402, 2001.
- [13] S. Periaswamy and H. Farid, *Medical image registration with partial data*, Med. Image Anal., vol. 10, no. 3, pp. 452-464, 2006.
- [14] B. Yeo, M. Sabuncu, R. Desikan, B. Fischl, P. Golland, *Effects of registration regularization and atlas sharpness on segmentation accuracy*, Med. Image Anal. vol. 12, no. 5, pp. 603-615, 2008.
- [15] R. Stefanescu, O. Commowick, G. Malandain, P.-Y. Bondiau, N. Ayache, and X. Pennec, *Non-rigid Atlas to Subject Registration with Pathologies for Conformal Brain Radiotherapy*, MICCAI 2004, LNCS 3216, pp. 704-711, 2004.
- [16] N. Cahill, J. Noble, D. Hawkes, *A demons algorithm for image registration with locally adaptive regularization*, In MICCAI2009. LNCS 5761, pp. 574-581. Springer, 2009.
- [17] P. Risholm, E. Samset, I. Talos, W. Wells, *A non-rigid registration framework that accommodates resection and retraction*, In IPMI2009. LNCS 5636, pp. 447-458. Springer, 2009.
- [18] L. Tang, G. Hamarneh, R. Abugharbieh, *Reliability-driven, spatially-adaptive regularization for deformable registration*, In WBIR2010. LNCS 6204, pp. 173-185. Springer, 2010.
- [19] H. Lamecker and X. Pennec, *Atlas to Image-with-Tumor Registration based on Demons and Deformation Inpainting*, In Proc. MICCAI Workshop on Computational Imaging Biomarkers for Tumors - From Qualitative to Quantitative (CIBT'2010), September 2010.
- [20] N. Chitphakdithai, K. P. Vives, and J. S. Duncan, *Registration of Brain Resection MRI with Intensity and Location Priors*, 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 1520-1523, 2011.
- [21] I. J. A. Simpson, J. A. Schnabel, A. R. Groves, J. L. R. Andersson, M. W. Woolrich, *Probabilistic inference of regularisation in non-rigid registration*, NeuroImage, vol. 59, no. 3, pp. 2438-2451, 2012.
- [22] M. Foskey, B. Davis, L. Goyal, S. Chang, E. Chaney, N. Strehl, S. Tomei, J. Rosenman, and S. Joshi, *Large deformation 3D image registration in image-guided radiation therapy*, Phys. Med. Biol., vol. 50, no. 24, pp. 5869-5892, 2005.
- [23] S. Gao, L. Zhang, H. Wang, R. de Crevoisier, D. D. Kuban, R. Mohan, and L. Dong, *A deformable image registration method to handle distended rectums in prostate cancer radiotherapy*, Medical Physics, vol. 33, no. 9, pp. 3304-3312, 2006.

- [24] M. Bach Cuadra, M. De Craene, V. Duay, B. Macq, C. Pollo, and J.-Ph. Thiran, *Dense deformation field estimation for atlas-based segmentation of pathological MR brain images*, Computer methods and programs in biomedicine, vol. 84, no. 2-3, pp. 66-75, 2006.
- [25] E. I. Zacharaki, C. S. Hoge, D. Shen, G. Biros, C. Davatzikos, *Non-diffeomorphic registration of brain tumor images by simulating tissue loss and tumor growth*, Neuroimage, vol. 46, no. 3, pp. 762-774, 2009.
- [26] A. Mang, S. Becker, A. Toma, and T. Buzug, *Coupling tumor growth with brain deformation: a constrained parametric non-rigid registration problem*, Medical Imaging 2010: Image Processing, Proc. of SPIE, vol. 7623, pp. 76230C-1-12, 2010.
- [27] M. Sdika and D. Pelletier, *Nonrigid registration of multiple sclerosis brain images using lesion inpainting for morphometry or lesion mapping*, Human brain mapping, vol. 30, no. 4, pp. 1060-1067, 2009.
- [28] M. Brett, A. P. Leff, C. Rorden, and J. Ashburner, *Spatial Normalization of Brain Images with Focal Lesion Using Cost Function Masking*, NeuroImage, vol. 14, no. 2, pp. 486-500, 2001.
- [29] S. M. Andersen, S. Z. Rapcsak, P. M. Beeson, *Cost function masking during normalization of brains with focal lesions: Still a necessity*, NeuroImage, vol. 53, no. 1, pp. 78-84, 2010.
- [30] P. Ripolles, J. Marco-Pallares, R. de Diego-Balaguer, J. Miro, M. Falip, M. Juncadella, F. Rubio, A. Rodriguez-Fornells, *Analysis of automated methods for spatial normalization of lesioned brains*, NeuroImage, vol. 60, no. 2, pp. 1296-1306, 2012.
- [31] B. Qin, Z. Gu, X. Sun, Y. Lv, *Registration of Images With Outliers Using Joint Saliency Map*, IEEE Signal Proc. Lett., vol. 17, no. 1, pp. 91-94, 2010.
- [32] Z. Gu, B. Qin, *Multi-modal and Multi-temporal Image Registration in the Presence of Gross Outliers Using Feature Voxel-Weighted Normalized Mutual Information*, IEEE Nuclear Science Symposium Conference Record, vol. 6, pp. 3209-3212, 2006.
- [33] B. Qin, Z. Gu, *Robust adaptive non-rigid image registration based on joint salient point sets in the presence of tumor-like gross outliers*, Proc. SPIE 6833, Electronic Imaging and Multimedia Technology V, 683320, November 28, 2007.
- [34] J. Zhou, B. Qin, *Geometrical regularization of nonrigid registration using local anisotropic structure and joint saliency map*, Proc. SPIE 8009, Third International Conference on Digital Image Processing (ICDIP 2011), 800913, July 08, 2011.
- [35] V. Katkovnik, A. Foi, K. Egiazarian, J. Astola, *From Local Kernel to Nonlocal Multiple-Model Image Denoising*, Int. J. Comput. Vis., vol. 86, no. 1, pp. 1-32, 2010.
- [36] H. Knutsson and C.-F. Westin, *Normalized and differential convolution: Methods for interpolation and filtering of incomplete and uncertain data*, in Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition, Jun. 16-19, pp. 515-523, 1993.
- [37] E. Suarez, C.-F. Westin, E. Rovaris, and J. Ruiz-Alzola, *Nonrigid registration using regularized matching weighted by local structure*, Medical Image Computing and Computer-Assisted Intervention-MICCAI 2002, LNCS 2489, pp. 581-589, 2002.
- [38] E. Suarez-Santana, R. Nebot, C.-F. Westin, and J. Ruiz-Alzola, *Fast BlockMatching Registration with Entropy-based Similarity*, Insight Journal November, pp. 1-8, 2006.
- [39] E. Ardizzone, R. Gallea, O. Gambino, and R. Pirrone, *Multi-modal Image Registration Using Fuzzy Kernel Regression*, Proceedings of the 16th IEEE international conference on Image processing, ICIP'09, pp. 193-196, 2009.
- [40] B. Liu, J. Zhang, X. Liao, *Projection Kernel Regression for Image Registration and Fusion in Video-Based Criminal Investigation*, International Conference on Multimedia and Signal Processing (CMSP'11), pp. 348-352, 2011.
- [41] C. Xing and P. Qiu, *Intensity-Based Image Registration by Nonparametric Local Smoothing*, IEEE Trans. Pattern. Anal. Mach. Intell., vol. 33, no. 10, pp. 2081-2092, 2011.
- [42] Z. Zhou, B. Qin, *Robust Non-Rigid Image Registration Using Incomplete Information, Adaptive Normalized Convolution and Similarity Measure Combination*, IEEE Nuclear Science Symposium Conference Record (NSS'07), pp. 3197-3201, 2007.
- [43] A. Andronache, M. von Siebenthal, G. Szekely, Ph. Cattin, *Non-rigid registration of multi-modal images using both mutual information and cross-correlation*, Med. Image Anal., vol. 12, no. 1, pp. 3-15, 2007.
- [44] M. Nitzberg and T. Shiota, *Nonlinear image filtering with edge and corner enhancement*, IEEE Trans. Pattern. Anal. Mach. Intell., vol. 14, no. 8, pp. 826-833, 1992.
- [45] V. Katkovnik, K. Egiazarian, and J. Astola, *A Spatially Adaptive Nonparametric Regression Image Deblurring*, IEEE Trans. Imag. Proc., vol. 14, no. 10, pp. 1469-1478, 2005.
- [46] T. Q. Pham, L. J. van Vliet, and K. Schutte, *Robust fusion of irregularly sampled data using adaptive normalized convolution*, EURASIP J. APPL. SIG. P. vol. 2006, pp. 1-12, 2006.
- [47] H. Takeda, S. Farsiu, P. Milanfar, *Kernel Regression for Image Processing and Reconstruction*, IEEE Trans. Imag. Proc., vol. 16, no. 2, pp. 349-366, 2007.
- [48] C. F. Westin, *A Tensor Framework for Multidimensional Signal Processing*, Linkping University, Sweden, 1994. ISBN 91-7871-421-4.
- [49] J. Weickert, *Anisotropic Diffusion in Image Processing*, B.G. Teubner Stuttgart, 1998.
- [50] A. Borji, L. Itti, *State-of-the-Art in Visual Attention Modeling*, IEEE Trans. Pattern. Anal. Mach. Intell., 35(1): 185-207, 2013.
- [51] H. Zhang, P. A. Yushkevich, J. C. Gee, *Registration of Diffusion Tensor Images*, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04), 2004.
- [52] H. Zhang, P. A. Yushkevich, D. C. Alexander, J. C. Gee, *Deformable registration of diffusion tensor MR images with explicit orientation optimization*, Med. Image Anal., vol. 10, no. 5, pp. 764-785, 2006.
- [53] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, J. C. Gee, *A reproducible evaluation of ANTs similarity metric performance in brain image registration*, NeuroImage, vol. 54, no. 3, pp. 2033-2044, 2011.
- [54] T. Vercauteren, X. Pennec, A. Perchant, N. Ayache, *Diffeomorphic demons: Efficient non-parametric image registration*, NeuroImage, vol. 45, pp. S61-S72, 2009.
- [55] T. Rohlfing, *Image Similarity and Tissue Overlaps as Surrogates for Image Registration Accuracy: Widely Used but Unreliable*, IEEE Trans. Med. Imaging, vol. 31, no. 2, pp. 153-163, 2012.
- [56] K. Murphy, B. van Ginneken, J. M. Reinhardt, S. Kabus, K. Ding, X. Deng, and et al., *Evaluation of Methods for Pulmonary Image Registration: The EMPIRE10 Study*, IEEE Trans. Med. Imaging, vol. 30, no. 11, pp. 1901-1920, 2011.
- [57] K. N. Chaudhury, S. Sanyal, *Improvements on "Fast Space-Variant Elliptical Filtering Using Box Splines"*, IEEE Trans. Image Proc., 21(9): 3915-3923, 2012.
- [58] W. Van Hecke, A. Leemans, S. De Backer, et al., *Comparing isotropic and anisotropic smoothing for voxel based DTI analyses: A simulation study*, Human brain mapping, vol. 31, no. 1, pp. 98-114, 2010.